US Macro Data Forecasting Report









This is a report on analyzing and forecasting the US macro data using **Recurrent Neural N Convolutional Neural Network** (CNNs) and **Generative Adversarial Net** report is:

Part I. Statistical analysis

- · Basic manipulation
- · Correlation analysis
- Time series analysis with ARIMA

Part II. Deep learning models

- Basic model: single-step, single-feature forecasting with LSTM
- · Generalized model: multi-step, multi-feature forcasting with LSTM
- Advanced model: Generative Adversarial Network (GAN) with RNN and CNN.

Part III. Conclusions and Next steps

- Conclusions
- Next steps

Introduction

1. The Notebook

Follow the notebook, we can recreate all the results, notice that

- Upload the USMacroData.xls file to the root folder on google colab.
- To navigate better, use the table of contents bottom on the upper-left sidebar.
- For clarity, all code cells are hiden, double click on the cell to get thε
- Change the parameters as indicated in the comments to create more custom outputs.
- All source code can also be found in the project file folder

2. The US Macro dataset

This report uses a US Macro Dataset provided by the ADP.

Before analyzing the data with codes, we have the following observations.

• This dataset contains 6 different features (the **Inflation**, **Wage**, **Unemployment**, **InterstRate**) about the macro economy of the US.

- Data were collected every 1 month, beginning in 1965-01-01 to 2015-12-01.
- In total, we have **612 rows** (month) and **6 columns** (features).

Part I.1 Basic manipulation

Code and examples

basic.py

read the file and show the head

₽		Month	Inflation	Wage	Unemployment	Consumption	Investment	InterestRat
	0	1965-01-01	1.557632	3.200000	4.9	6.972061	12.3	3.9
	1	1965-02-01	1.557632	3.600000	5.1	7.811330	13.2	3.9
	2	1965-03-01	1.242236	4.000000	4.7	7.828032	18.7	4.(
	3	1965-04-01	1.552795	3.585657	4.8	8.477938	9.8	4.(
	4	1965-05-01	1.552795	3.968254	4.6	7.139364	10.2	4.1

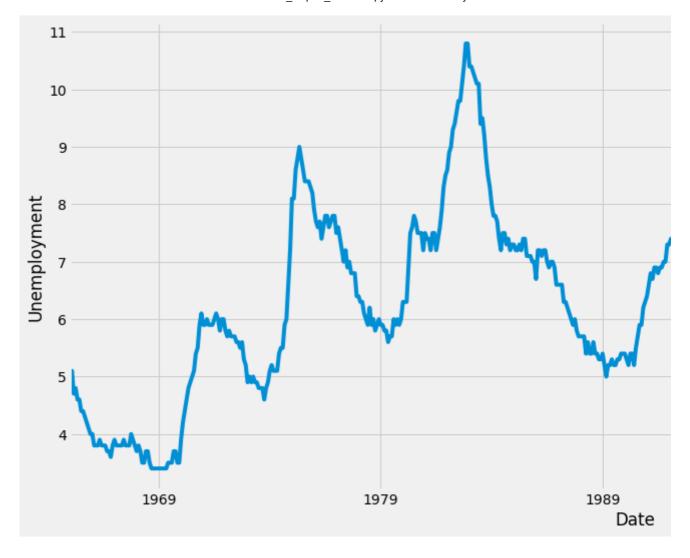
Basic checks: find null values and fill, set index, etc.

Null values summary:

Inflation	0	
Wage	0	
Unemployment	0	
Consumption	0	
Investment	0	
InterestRate		
dtype: int64		

	Inflation	Wage	Unemployment	Consumption	Investment	InterestRate
Month						
1965-01-01	1.557632	3.200000	4.9	6.972061	12.3	3.90
1965-02-01	1.557632	3.600000	5.1	7.811330	13.2	3.98
1965-03-01	1.242236	4.000000	4.7	7.828032	18.7	4.04
1965-04-01	1.552795	3.585657	4.8	8.477938	9.8	4.09
1965-05-01	1.552795	3.968254	4.6	7.139364	10.2	4.10

Example: plot the "Inflation" column



Data Analysis

As a high level overview, some distinguishable patterns appear when we plot the data:

- In the 80's (1979-1989), all features experienced some drastic change
- The time-series has **seasonality pattern**, for example, **Unemployment** has **long** goes through 1 or 2 major up and downs. We will examine the seasonality more carefully in

Part I.2 Correlation analysis

Though it's indicated that there's no obvious correlation among the 6 features, we compute severa **Naive correlation**, **Pearson correlation**, **local Pearson correlation**, **instan** and related statistics in order to

- Test the validity of the assumption (i.e. no two features are apprantly correlated).
- Chose source and target features for later model builds.

By doing so, we can get more understanding about the 'quality' and 'inner relations' of the data. If a explanatory power to the feature that we want to predict (e.g. "Inflation"), then there is no need for learning models. On the other hand, if one feature has higher-than-random correlations to another

the feature and the other as the target. In this case, to determine which feature leads, the Dynamic time wrapping.

Code and Examples

correlation.py

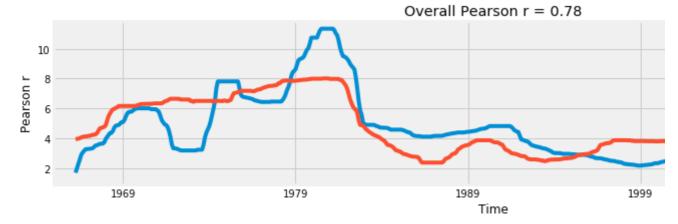
Requirement already satisfied: dtw in /usr/local/lib/python3.6/dist-packages (1.4.0)
Requirement already satisfied: scipy in /usr/local/lib/python3.6/dist-packages (from Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from

Example: Naive correlation.

₽		Inflation	Wage	Unemployment	Consumption	Investment	InterestR
	Inflation	1.000000	0.778155	0.191886	0.617820	-0.341421	0.773
	Wage	0.778155	1.000000	-0.068529	0.703745	-0.125412	0.647
	Unemployment	0.191886	-0.068529	1.000000	-0.097183	-0.038286	-0.027
	Consumption	0.617820	0.703745	-0.097183	1.000000	0.203165	0.655
	Investment	-0.341421	-0.125412	-0.038286	0.203165	1.000000	-0.234
	InterestRate	0.773616	0.647482	-0.027809	0.655305	-0.234573	1.000

Example: Pearson correlation

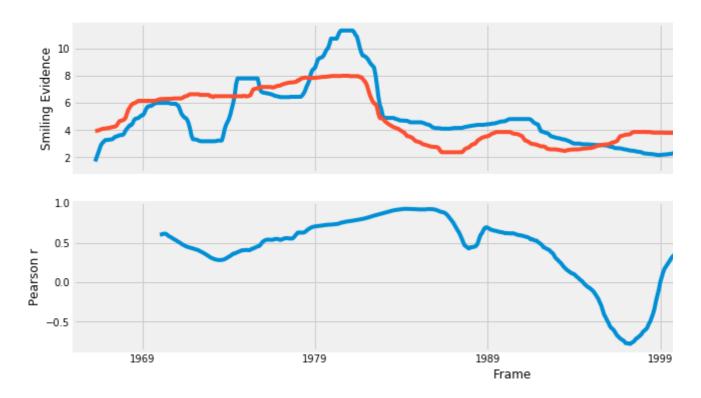
Pandas computed Pearson r: 0.7781551675438367
Scipy computed Pearson r: 0.7781551675438365 and p-value: 2.53137614903759e-125



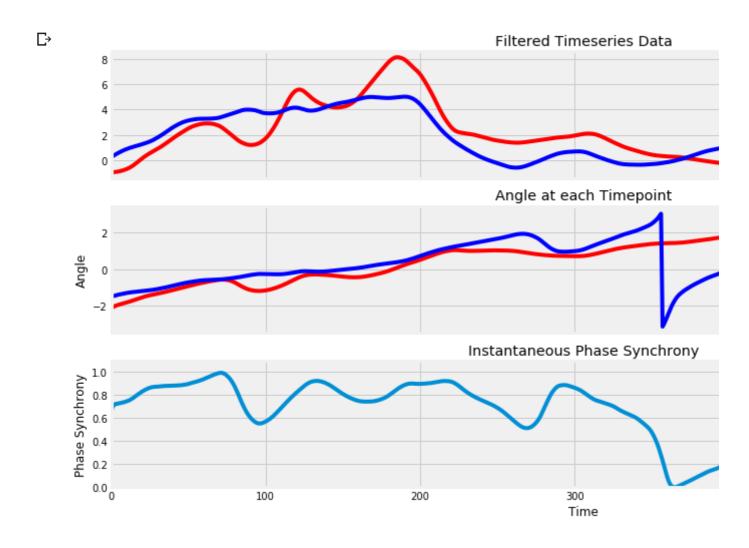
Example: local Pearson correlation

С→

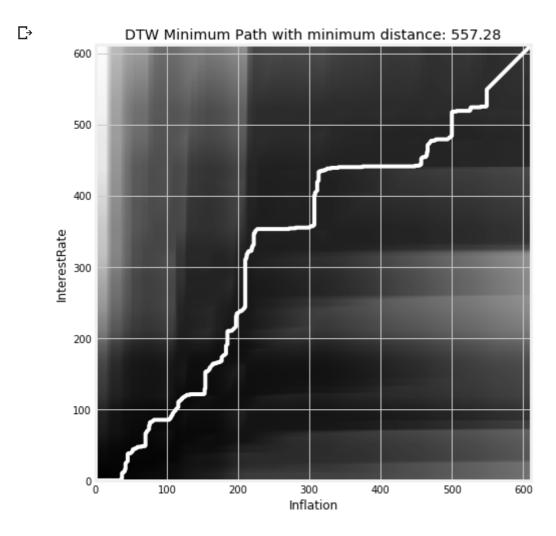
Smiling data and rolling window correlation



Example: instantaneous phase synchronization



Example: dynamic time wraping



Data analysis

Inspecting the correlations from different angles, we find

- Inflation and Wage have the highest correlation, 0.778155, among all the
- Inflation, Wage, Consumption and IntestRate show quite high positive correlation, and low n Unemployment and Investment.
- Most features slightly leads the Inflation feature.
- For the first 30 years, certain feature pairs show **high instantaneous phase synch**1

We conclude that

- The assumption that no two features have apparent correlation is w
- It's reasonable to
 use Inflation as target and the other 5 features as source for forcasting

Part I.3 Time series analysis with ARIMA

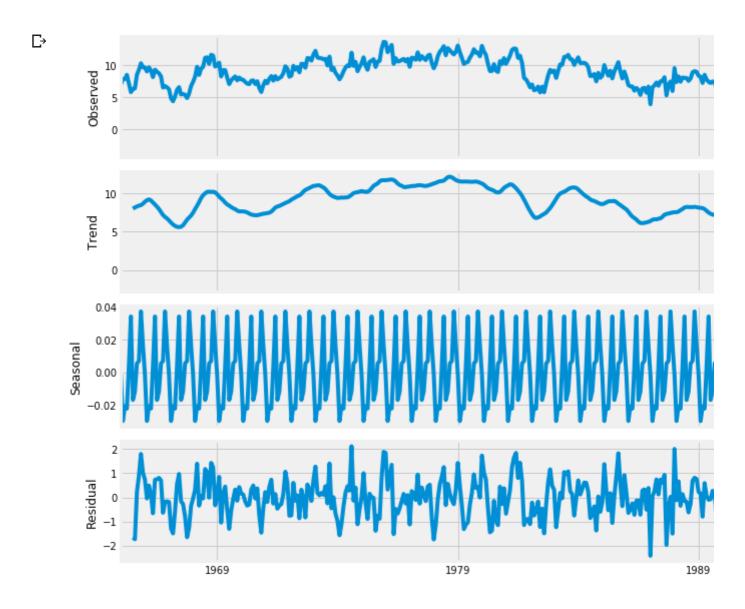
As we mentioned above, some remarkable patterns (e.g. seasonality pattern) naturally appear in c

- We visualize our data using **time-series decomposition** that allows us to decompos trend, seasonality, and noise.
- We train an ARIMA (Autoregressive Integrated Moving Average) m Inflation values. To get optimal output, we first
- Use **grid** search to get the optimal parameters for the ARIMA mode.
- We use ARIMA diagnostics to investigate any unusual behavior.

Code and examples

time_series.py

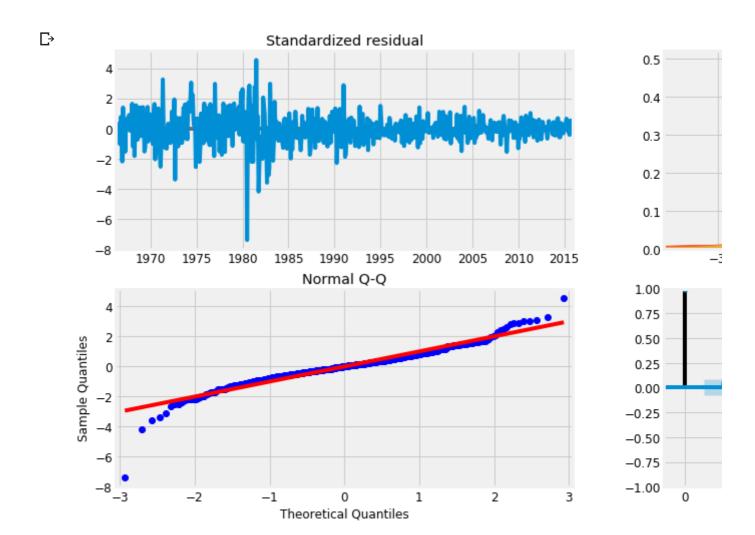
Example: decompose "Consumption" column into trend, seasonal and residua



Time series analysis with ARIMA Grid search for optimal ARIMA parameters

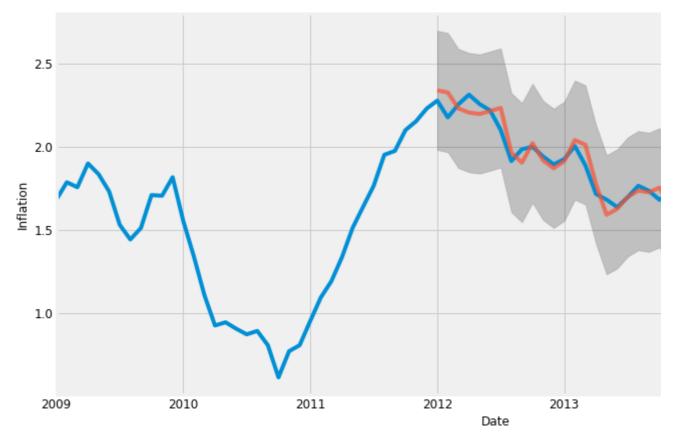
ARIMA training

ARIMA diadonostics



ARIMA predictions

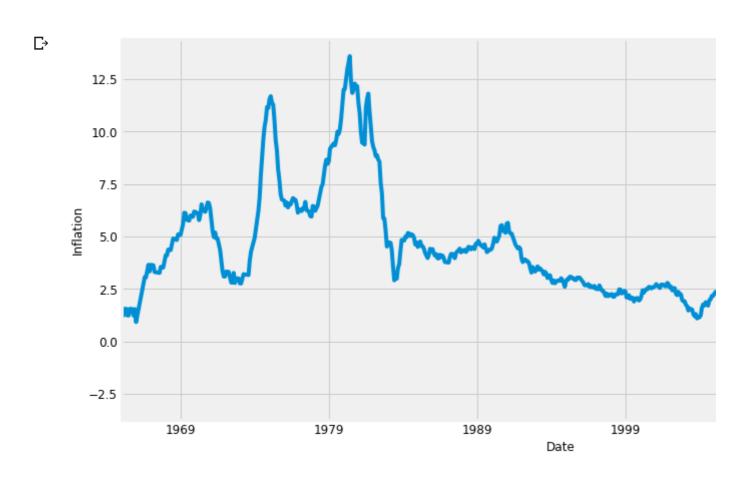
 \Box



The Mean Squared Error of our forecasts is 0.004963826636415743

The Root Mean Squared Error of our forecasts is 0.07

ARIMA forcasts



Data Analysis

- Components plot show the obvious seasonality, for example, in every 10 years, the "Inflatial a half-year seasonality."
- The optimal ARIMA parameters for "Inflation" are (1, 1, 1)x(0, 0, 1, 12)
- The ARIMA diagonostics show that the **noise distribution is narrower than the**
- The one-step ahead forcast captures the overall trend well.
- As we forecast further out into the future, we becomes less confident in our values. This is r by our model, which grow larger as we move further out into the future.

Part II.1 Basic model: single-step, single-feature fo

Recurrent Neural Networks (RNNs) are good fits for time-series analysis because f designed to capture patterns developing through time.

However, vanilla RNNs have a major disadvantage---the vanishing gradient problem---"the changes so small, making the network unable to converge to a optimal solution.

LSTM (**Long-Short Term Memory**) is a variation of vanilla RNNS,it overcomes the variable problem by clipping gradients if they exceed some constant bounds.

In this section, we will

- Process the data to fit the LSTM model
- Build and train the LSTM model for single-step, single-feature pred tomorrow value with only today's values of the other 5 features).

imports

Data preparation

Build and train the LSTM model

Make sure data forms are correct

LSTM with SGD, RMSprop, Adam optimizers, epochs = 100

```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/optimizers.py:75
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow_core/pythor
Instructions for updating:
Use tf.where in 2.0, which has the same broadcast rule as np.where
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
Train on 427 samples, validate on 184 samples
Epoch 1/100
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/keras/backend/tensorfl
Epoch 2/100
Epoch 3/100
427/427 [============== ] - 1s 3ms/step - loss: 1.8379 - acc: 0.0000e+
Epoch 4/100
427/427 [================ ] - 1s 3ms/step - loss: 1.4932 - acc: 0.0000e+
Epoch 5/100
Epoch 6/100
Epoch 7/100
427/427 [============== ] - 1s 3ms/step - loss: 1.0970 - acc: 0.0000e+
Epoch 8/100
Epoch 9/100
Epoch 10/100
427/427 [============= ] - 1s 3ms/step - loss: 1.0670 - acc: 0.0000e+
Epoch 11/100
427/427 [============== ] - 1s 3ms/step - loss: 1.0320 - acc: 0.0000e+
Epoch 12/100
Epoch 13/100
427/427 [============== ] - 1s 3ms/step - loss: 0.9960 - acc: 0.0000e+
Epoch 14/100
Epoch 15/100
Epoch 16/100
Epoch 17/100
```

```
Epoch 18/100
Epoch 19/100
427/427 [============= ] - 1s 3ms/step - loss: 0.9319 - acc: 0.0000e+
Epoch 20/100
Epoch 21/100
Epoch 22/100
Epoch 23/100
Epoch 24/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.8879 - acc: 0.0000e+
Epoch 25/100
Epoch 26/100
Epoch 27/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.8588 - acc: 0.0000e+
Epoch 28/100
427/427 [============ ] - 1s 3ms/step - loss: 0.8365 - acc: 0.0000e+
Epoch 29/100
Epoch 30/100
Epoch 31/100
427/427 [============== ] - 1s 3ms/step - loss: 0.7964 - acc: 0.0000e+
Epoch 32/100
Epoch 33/100
427/427 [================ ] - 1s 3ms/step - loss: 0.8114 - acc: 0.0000e+
Epoch 34/100
Epoch 35/100
Epoch 36/100
Epoch 37/100
Epoch 38/100
Epoch 39/100
Epoch 40/100
427/427 [============== ] - 1s 3ms/step - loss: 0.7406 - acc: 0.0000e+
Epoch 41/100
Epoch 42/100
Epoch 43/100
Epoch 44/100
Epoch 45/100
Epoch 46/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.7355 - acc: 0.0000e+
Epoch 47/100
Fnoch 48/100
```

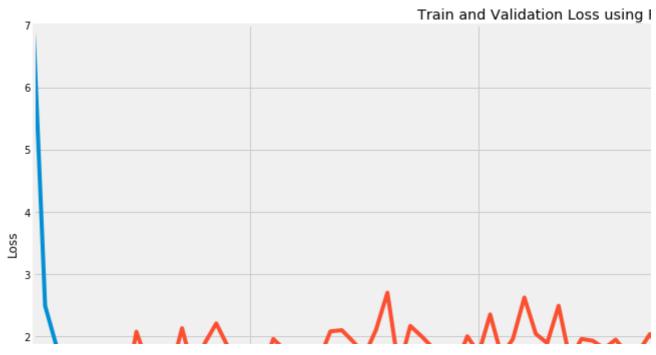
```
LPUCH -10, 100
427/427 [============ ] - 1s 3ms/step - loss: 0.7791 - acc: 0.0000e+
Epoch 49/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.7407 - acc: 0.0000e+
Epoch 50/100
Epoch 51/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.7301 - acc: 0.0000e+
Epoch 52/100
Epoch 53/100
Epoch 54/100
427/427 [============= ] - 1s 3ms/step - loss: 0.7170 - acc: 0.0000e+
Epoch 55/100
Epoch 56/100
Epoch 57/100
Epoch 58/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.7016 - acc: 0.0000e+
Epoch 59/100
Epoch 60/100
427/427 [============== ] - 1s 3ms/step - loss: 0.7264 - acc: 0.0000e+
Epoch 61/100
427/427 [============= ] - 1s 3ms/step - loss: 0.6682 - acc: 0.0000e+
Epoch 62/100
Epoch 63/100
427/427 [================ ] - 1s 3ms/step - loss: 0.6951 - acc: 0.0000e+
Epoch 64/100
Epoch 65/100
Epoch 66/100
Epoch 67/100
Epoch 68/100
Epoch 69/100
427/427 [================ ] - 1s 3ms/step - loss: 0.6832 - acc: 0.0000e+
Epoch 70/100
Epoch 71/100
Epoch 72/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.6946 - acc: 0.0000e+
Epoch 73/100
Epoch 74/100
Epoch 75/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.6901 - acc: 0.0000e+
Epoch 76/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6839 - acc: 0.0000e+
Epoch 77/100
Epoch 78/100
```

```
Epoch 79/100
Epoch 80/100
Epoch 81/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6555 - acc: 0.0000e+
Epoch 82/100
Epoch 83/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6115 - acc: 0.0000e+
Epoch 84/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6476 - acc: 0.0000e+
Epoch 85/100
Epoch 86/100
Epoch 87/100
427/427 [================ ] - 1s 3ms/step - loss: 0.6399 - acc: 0.0000e+
Epoch 88/100
Epoch 89/100
Epoch 90/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.6269 - acc: 0.0000e+
Epoch 91/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6289 - acc: 0.0000e+
Epoch 92/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6331 - acc: 0.0000e+
Epoch 93/100
Epoch 94/100
Epoch 95/100
427/427 [=============== ] - 1s 3ms/step - loss: 0.6257 - acc: 0.0000e+
Epoch 96/100
427/427 [============== ] - 1s 3ms/step - loss: 0.6182 - acc: 0.0000e+
Epoch 97/100
Epoch 98/100
427/427 [============= ] - 1s 3ms/step - loss: 0.5794 - acc: 0.0000e+
Epoch 99/100
Epoch 100/100
```

Plot result

Plot result

С→

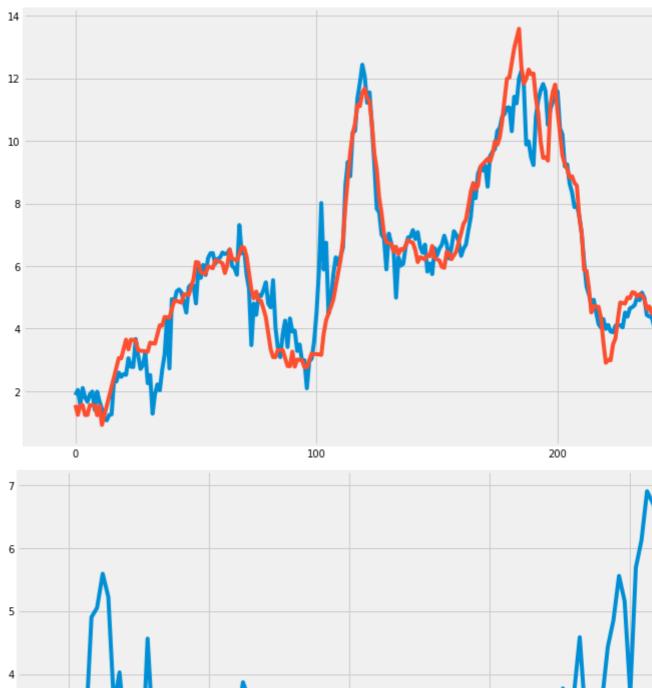


Plot predictions

Plot predictions

₽

Train Score: 0.71 RMSE Test Score: 1.53 RMSE



Data Analysis

We only trained the model for 100 epochs, feel free to modify it to any number as long as we have results we find during the experiments

- LSTM with Adam or RMSprop optimizers work better than the SGD optimizer in this project.
- Each model fits the training dataset very well.
- · The prediction captures the range and characteristics of the real dat
- The model doesn't predict the rapid increasing near the 100th test ${\bf d}$

Part II.2 Generalized model: multi-step, multi-fea

We build a multi-step, multi-feature LSTM model in this section. That means we can use several-d features in the future.

For example, we can use last 12-month's data of Wage, Consumption, In . In this section, we

- Process the data to fit the requirements of all possible multi-step, multi-feature prediction ta
- · We modify the LSTM model accordingly.
- Plot the 3-month prediction for Inflation and Unemployment with last 12-month's data of Wa

Data preparation

Make the data forms are all correct

```
(418, 12, 4)
(418, 6)
(180, 12, 4)
(418, 6)
```

Scaling, vectorize and de_vectorize

Multi-step LSTM model, change the input_shape and Dense layer parameter to

Train the model. Change the optimizer parameter to use other optimizers, e.g.

С→

```
Train on 418 samples, validate on 180 samples
Epoch 1/200
Epoch 2/200
Epoch 3/200
Epoch 4/200
Epoch 5/200
Epoch 6/200
Epoch 7/200
Epoch 8/200
Epoch 9/200
Epoch 10/200
Epoch 11/200
Epoch 12/200
Epoch 13/200
Epoch 14/200
Epoch 15/200
Epoch 16/200
Epoch 17/200
Epoch 18/200
Epoch 19/200
Epoch 20/200
Epoch 21/200
Epoch 22/200
Epoch 23/200
Epoch 24/200
Epoch 25/200
Epoch 26/200
Epoch 27/200
Epoch 28/200
Epoch 29/200
Epoch 30/200
```

```
Epoch 31/200
Epoch 32/200
Epoch 33/200
Epoch 34/200
Epoch 35/200
Epoch 36/200
Epoch 37/200
Epoch 38/200
Epoch 39/200
Epoch 40/200
Epoch 41/200
Epoch 42/200
Epoch 43/200
Epoch 44/200
Epoch 45/200
Epoch 46/200
Epoch 47/200
Epoch 48/200
Epoch 49/200
Epoch 50/200
Epoch 51/200
Epoch 52/200
Epoch 53/200
Epoch 54/200
Epoch 55/200
Epoch 56/200
Epoch 57/200
Epoch 58/200
Epoch 59/200
Epoch 60/200
Epoch 61/200
```

```
------
Epoch 62/200
Epoch 63/200
Epoch 64/200
Epoch 65/200
Epoch 66/200
Epoch 67/200
Epoch 68/200
Epoch 69/200
Epoch 70/200
Epoch 71/200
Epoch 72/200
Epoch 73/200
Epoch 74/200
Epoch 75/200
Epoch 76/200
Epoch 77/200
Epoch 78/200
Epoch 79/200
Epoch 80/200
Epoch 81/200
Epoch 82/200
Epoch 83/200
Epoch 84/200
Epoch 85/200
Epoch 86/200
Epoch 87/200
Epoch 88/200
Epoch 89/200
Epoch 90/200
Epoch 91/200
Epoch 92/200
```

```
Epoch 93/200
Epoch 94/200
Epoch 95/200
Epoch 96/200
Epoch 97/200
Epoch 98/200
Epoch 99/200
Epoch 100/200
Epoch 101/200
Epoch 102/200
Epoch 103/200
Epoch 104/200
Epoch 105/200
Epoch 106/200
Epoch 107/200
Epoch 108/200
Epoch 109/200
Epoch 110/200
Epoch 111/200
Epoch 112/200
Epoch 113/200
Epoch 114/200
Epoch 115/200
Epoch 116/200
Epoch 117/200
Epoch 118/200
Epoch 119/200
Epoch 120/200
Epoch 121/200
Epoch 122/200
Epoch 123/200
```

```
Epoch 124/200
Epoch 125/200
Epoch 126/200
Epoch 127/200
Epoch 128/200
Epoch 129/200
Epoch 130/200
Epoch 131/200
Epoch 132/200
Epoch 133/200
Epoch 134/200
Epoch 135/200
Epoch 136/200
Epoch 137/200
Epoch 138/200
Epoch 139/200
Epoch 140/200
Epoch 141/200
Epoch 142/200
Epoch 143/200
Epoch 144/200
Epoch 145/200
Epoch 146/200
Epoch 147/200
Epoch 148/200
Epoch 149/200
Epoch 150/200
Epoch 151/200
Epoch 152/200
Epoch 153/200
```

```
Epoch 154/200
Epoch 155/200
Epoch 156/200
Epoch 157/200
Epoch 158/200
Epoch 159/200
Epoch 160/200
Epoch 161/200
Epoch 162/200
Epoch 163/200
Epoch 164/200
Epoch 165/200
Epoch 166/200
Epoch 167/200
Epoch 168/200
Epoch 169/200
Epoch 170/200
Epoch 171/200
Epoch 172/200
Epoch 173/200
Epoch 174/200
Epoch 175/200
Epoch 176/200
Epoch 177/200
Epoch 178/200
Epoch 179/200
Epoch 180/200
Epoch 181/200
Epoch 182/200
Epoch 183/200
Epoch 184/200
```